Machine Learning Applications for Inferring Properties of Asteroids Impacting Earth

Ana Maria Tarano
December 2nd, 2019
Space Environment and Satellite Systems

Overview

As a body enters the atmosphere, its mechanical energy transforms into heat, light, and ionization.

However, these energy signatures do not provide direct measurements of the meteoroid's properties, which are critical to establish reliable asteroid impact risk assessments.

This presentation describes methods to infer meteoroid parameters based on the observed light curve from a meteoroid's flight across the atmosphere.

Table of Contents

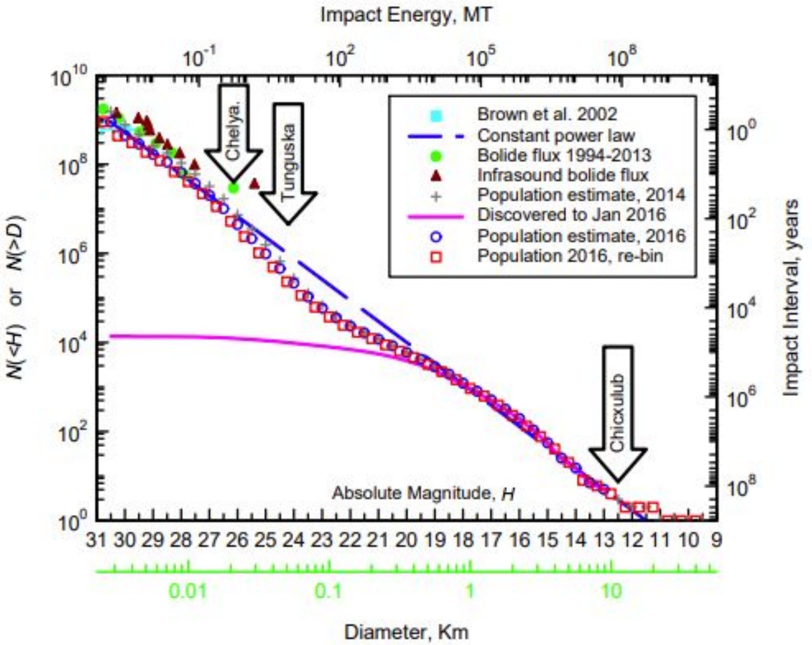
Introduction

01

Objects greater than **1 km** are assumed to lead to global damage, including potential human extinction.

Focus on search of potentially hazardous objects are on > **140 m** diameter.

The smaller, more frequent asteroids are not as frequently discovered but are still dangerous (e.g. Chelyabinsk **~20 m**).



Source: SDT NASA, 2017

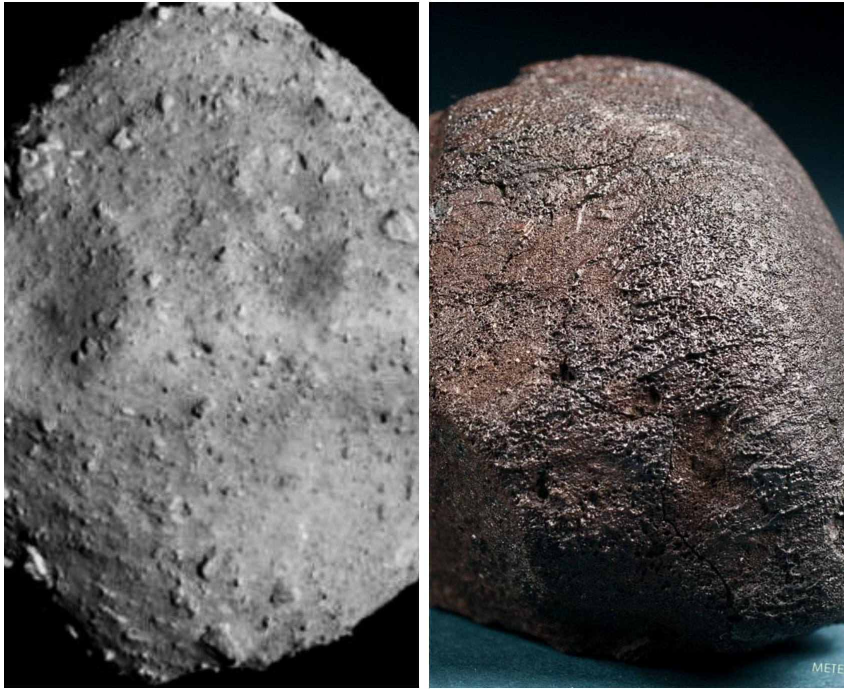
Introduction

02

Every day, 80 to 100 tons of material falls upon Earth from space in the form of dust and small meteorites.

Asteroid missions are expensive and meteorites are rare.

Earth's atmosphere is a naturally occurring laboratory that should be used to better understand asteroids, especially to better understand those that might pose a threat.



Source: NASA

Introduction
03

Studying meteoroids entering the atmosphere is becoming more readily available through the presence of space assets, such as the Geostationary Lightning Mapper, and growing ground-based camera networks.

Other types of observational methods, such as seismic, infrasound or spectroscopy, don't map the flight of the bolides.

Pre-entry parameters, such as diameter, density, angle, velocity, and bulk strength, are critical for asteroid threat assessment.

The pre-entry parameters of impacting asteroids are not directly measured from energy deposition curves derived from optical sensors.

Physics-based models and uncertain mean values are used to infer unknown quantities from energy deposition curves when velocity and entry angle are known.

Introduction

04

Light curve data from asteroids entering the atmosphere is abundant but the asteroid's properties are not commonly directly observed so there is ample but incomplete and unlabeled data.

The meteor and asteroid communities rely heavily on modeling to infer properties from the data by reproducing the energy signatures that were observed.

We have previously used a genetic algorithm to reproduce the manual labor of curve matching to solve for model inputs using a semi-analytical fragment-cloud model (FCM).

We leverage an extended version of FCM to generate labeled data to train regression models in order to infer model inputs from observed cases.

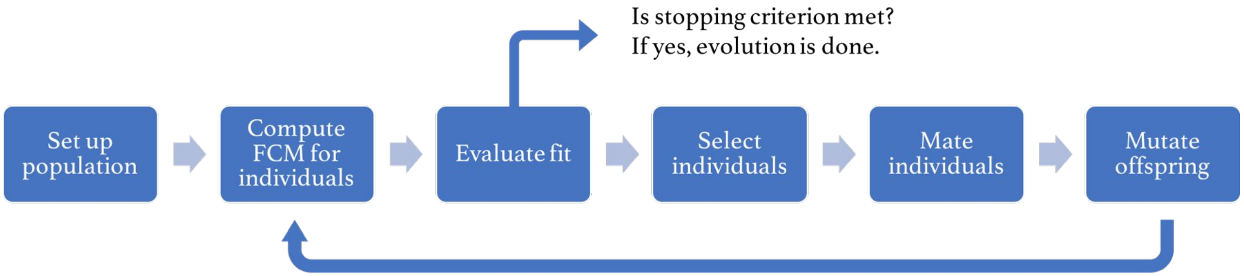
Study objective: Can our synthetically trained regression models be generalized to infer parameters from real fireballs?

Objectives

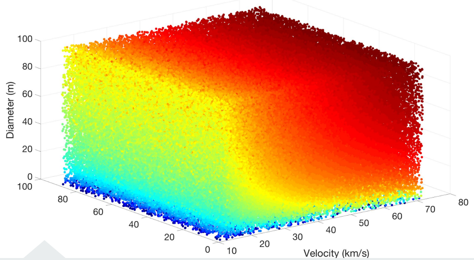
1. Infer pre-entry meteoroid parameters from measured meteor light curves.
2. Automate inference using supervised learning methods.
3. Identify the best performing tuning parameters for machine learning (hyperparameters).
4. Validate approach against well-documented but different test cases: Chelyabinsk, Lost City, Kosice, Tagish Lake, and Benešov meteors.

Review GA Methodology

Genetic algorithms (GAs) are iterative stochastic search-based optimizers based on evolution and genetics.



Review GA Methodology

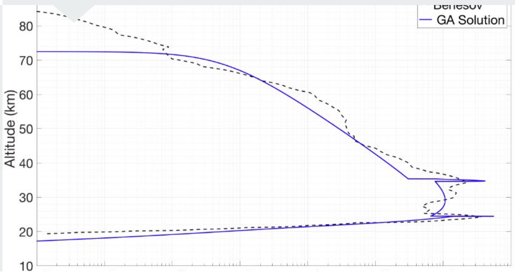


Develop GA with Synthetic Curves

Understand your search space without the limitations from the energy deposition model.

Test with Real Curves 15x

Understand strengths and challenges of reduced model to match observations.

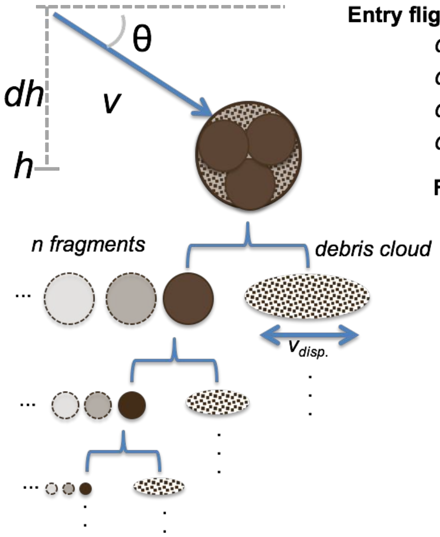


Parameter	Published Estimates	List of References
Mass (kg)	7E6 - 1.3E7	Wheeler et al. (2018);
Diameter (m)	15.2 - 24.4	Wheeler et al. (2017);
Density (g/cm^3)	2.2 - 3.3	Brown et al. (2016)
		Borovička (2015);
		Avramenko et al. (2014);
		Popova et al. (2013)

Compare Results with Literature

Understand if inference parallels other modelers for cases that have different features and meteorites.

GA Fragment-Cloud Model (FCM)



Entry flight: integrates meteor equations of motion and ablation

$$dm/dt = -0.5\rho_{air}v^3A\sigma$$
$$dv/dt = \rho_{air}v^2AC_D/m - g\sin\theta$$
$$d\theta/dt = (v/(R_E+h) - g/v)\cos\theta$$
$$dh/dt = v\sin\theta$$

Fragmentation when pressure > strength

$$\rho_{air}v^2 > strength$$

Each break yields:

- Multiple independent, identical fragments
- Debris cloud of specified mass fraction

Fragment strengths increase with decreased size

$$S_2 = S_1(m_1/m_2)^\alpha$$

Clouds broaden and slow under common bow shock

$$v_{dispersion} = v_{cloud}(3.5\rho_{air}A/\rho_{cloud})^{1/2}$$

Energy deposition computed as change in total KE of all fragments/clouds as a function of altitude.

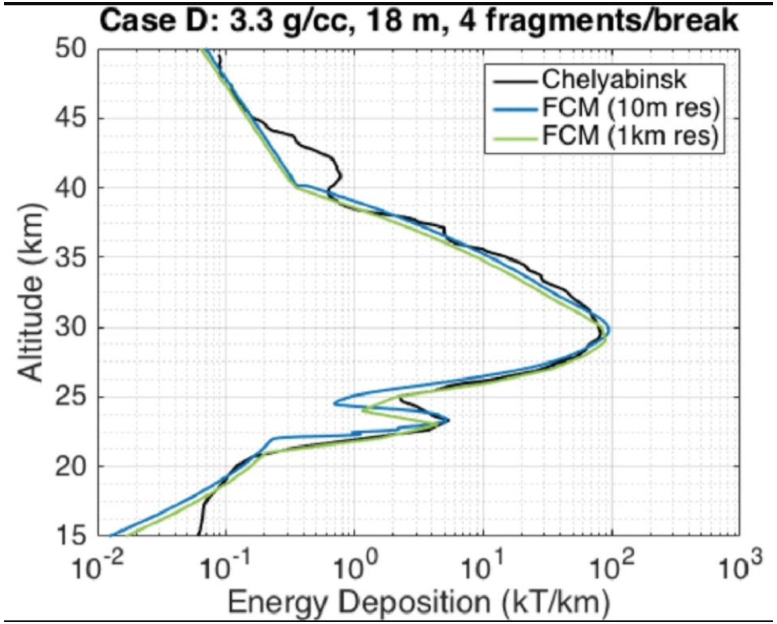
FCM Parameters

ρ_{air}: atmospheric density
m: mass
v : velocity
θ: entry angle with respect to horizontal
h : altitude
g: gravitational constant
C_D: drag coefficient
A: cross-sectional area of the bolide piece
σ: ablation parameter

S_2: post-break strength of child
S_1: pre-break strength of the parent fragment
m_2: post-break mass of child
m_1: pre-break mass of the parent fragment
α: strength scaling parameter

$v_{dispersion}$: cloud dispersion velocity
v_{cloud}: cloud velocity
ρ_{cloud} : cloud density

FCM Curve Matching



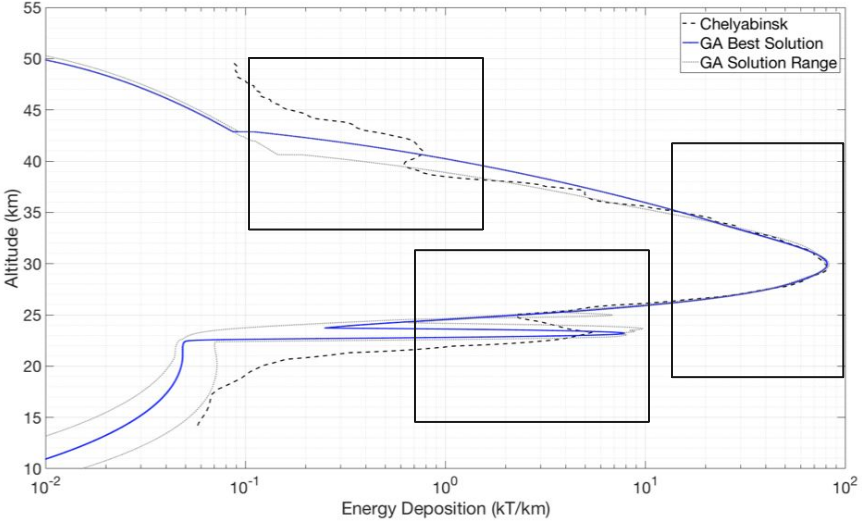
Modeling Assumptions

- σ is constant and the same for clouds and fragments.

- Body and fragments are spherical in shape and monolithic.

- Initial velocity and entry angle are known and accurate.

GA Review Chelyabinsk (single main flare)

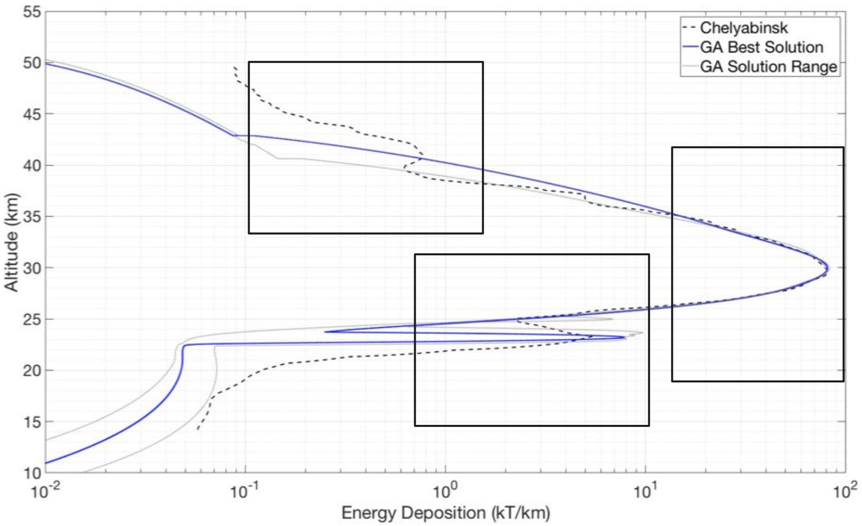


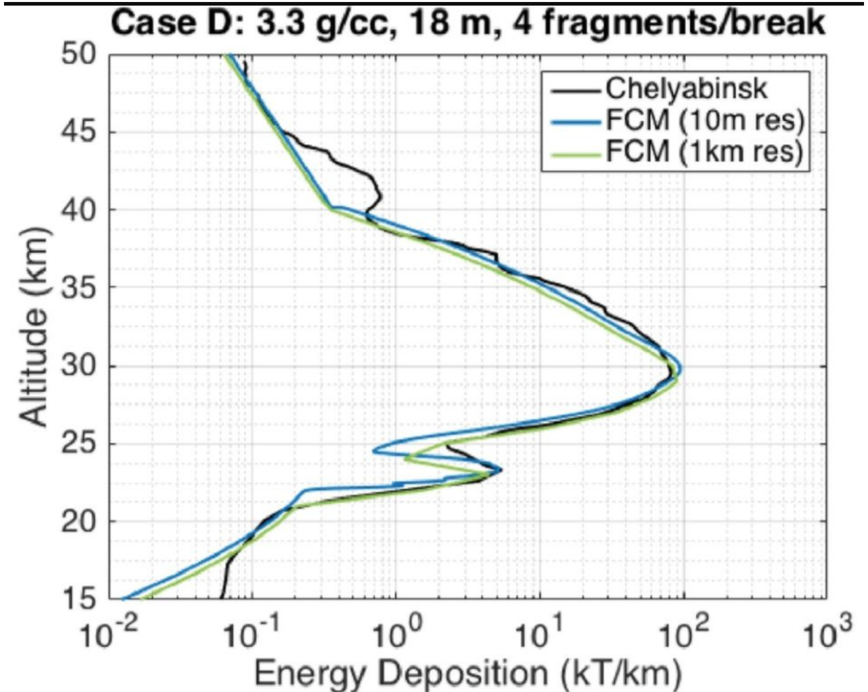
Chelyabinsk's solutions from GA consistently matched the main flare early in the evolution process.

The height and magnitude of second peak also consistently solved for. Modeling assumptions and objective function led to narrowness.

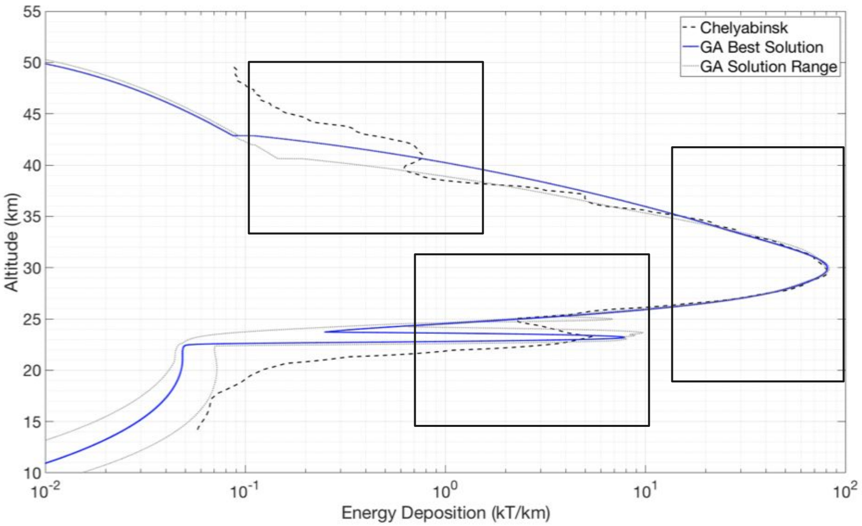
The top flare could not be modeled manually, requires non-monolithic body.

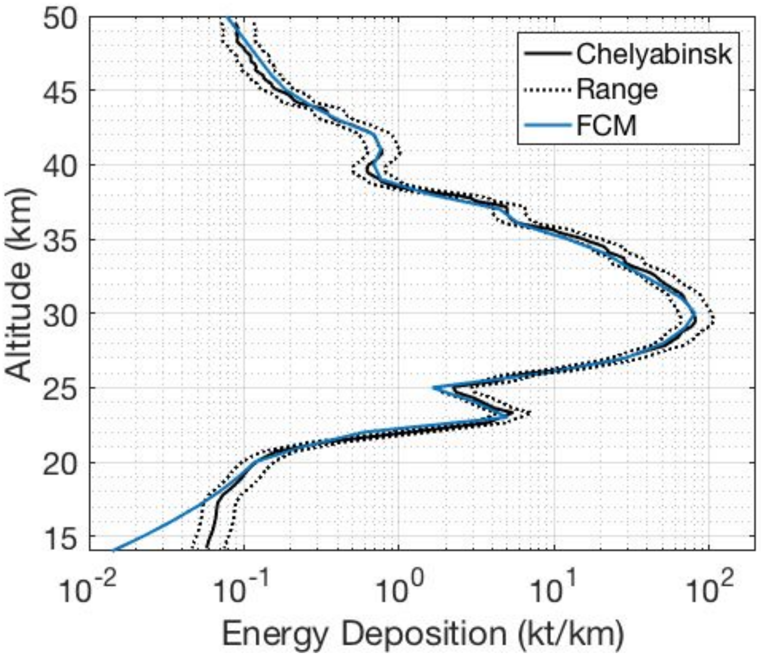
GA Review Chelyabinsk (single main flare)



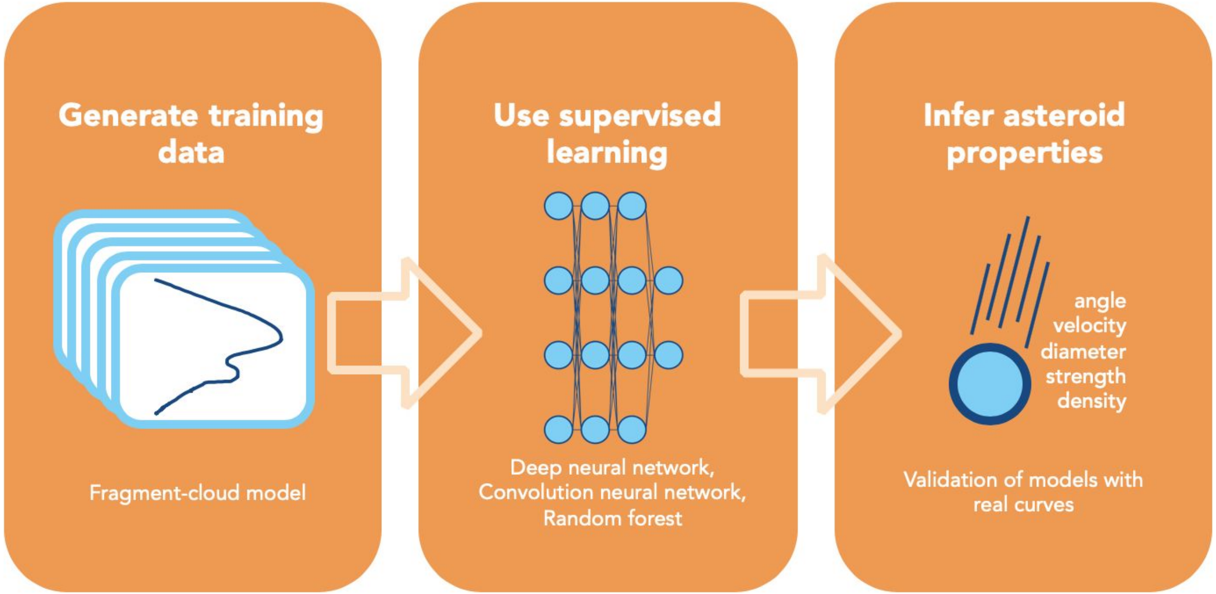


GA FCM Compared to FCM-RP Capabilities

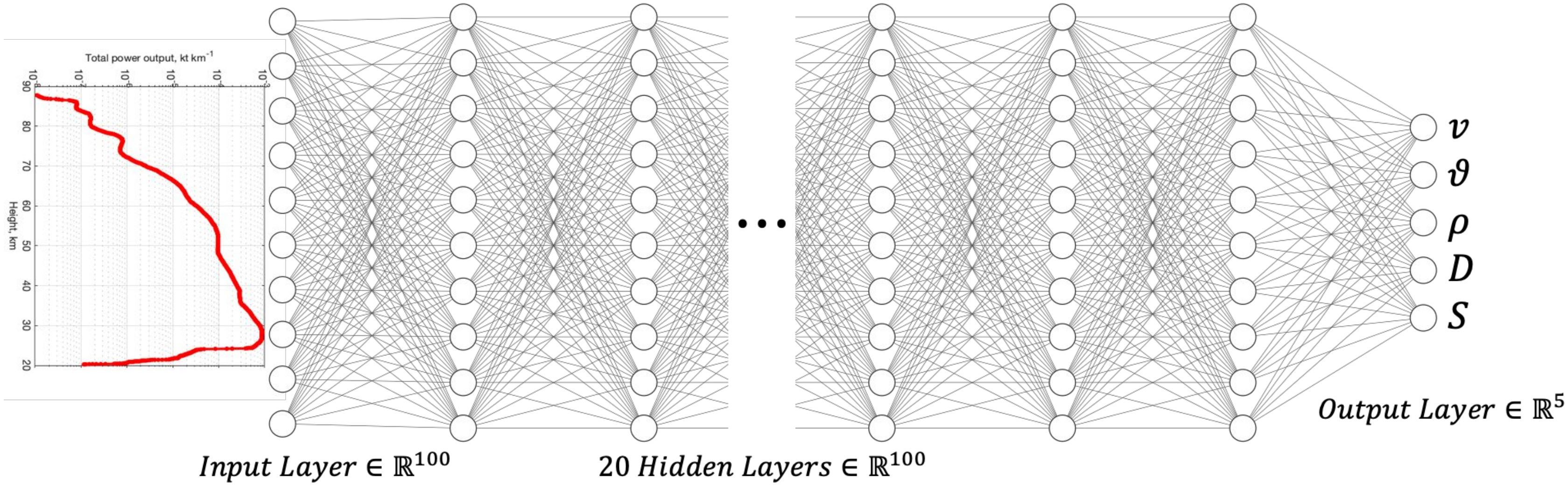




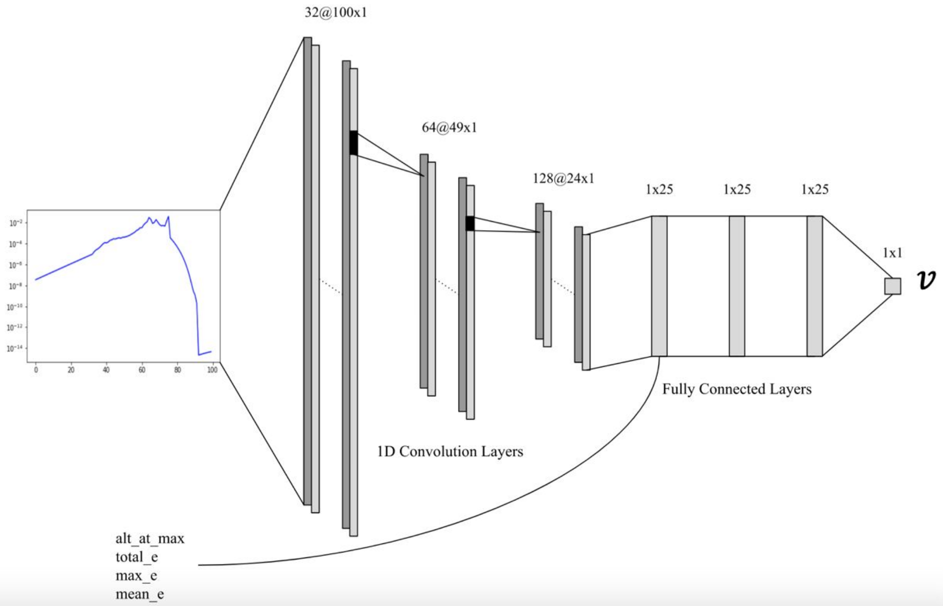
Regression Models Methodology Overview



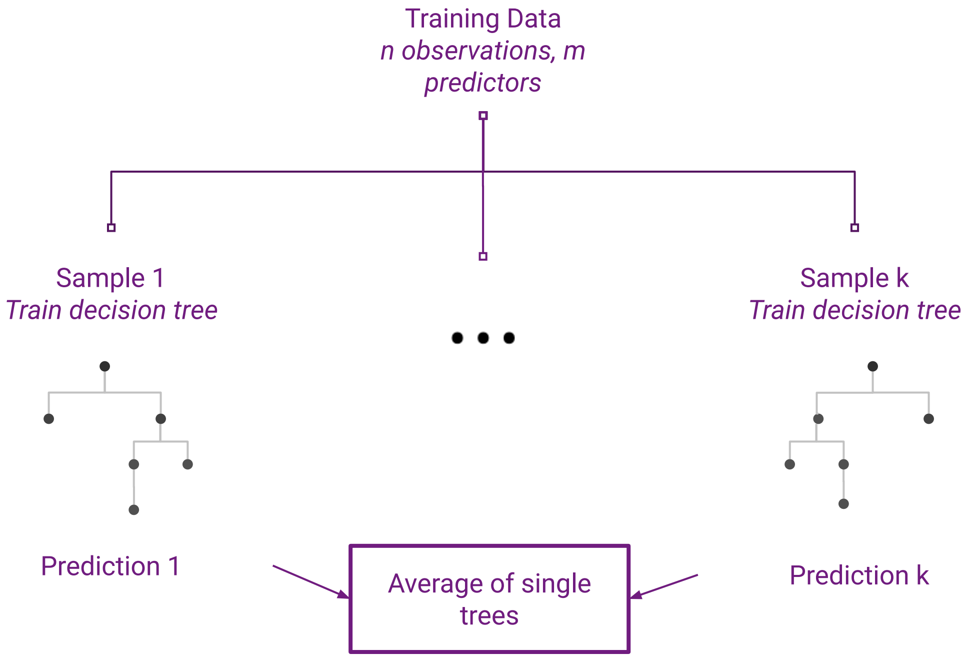
Deep Neural Network Model



Convolutional Neural Network

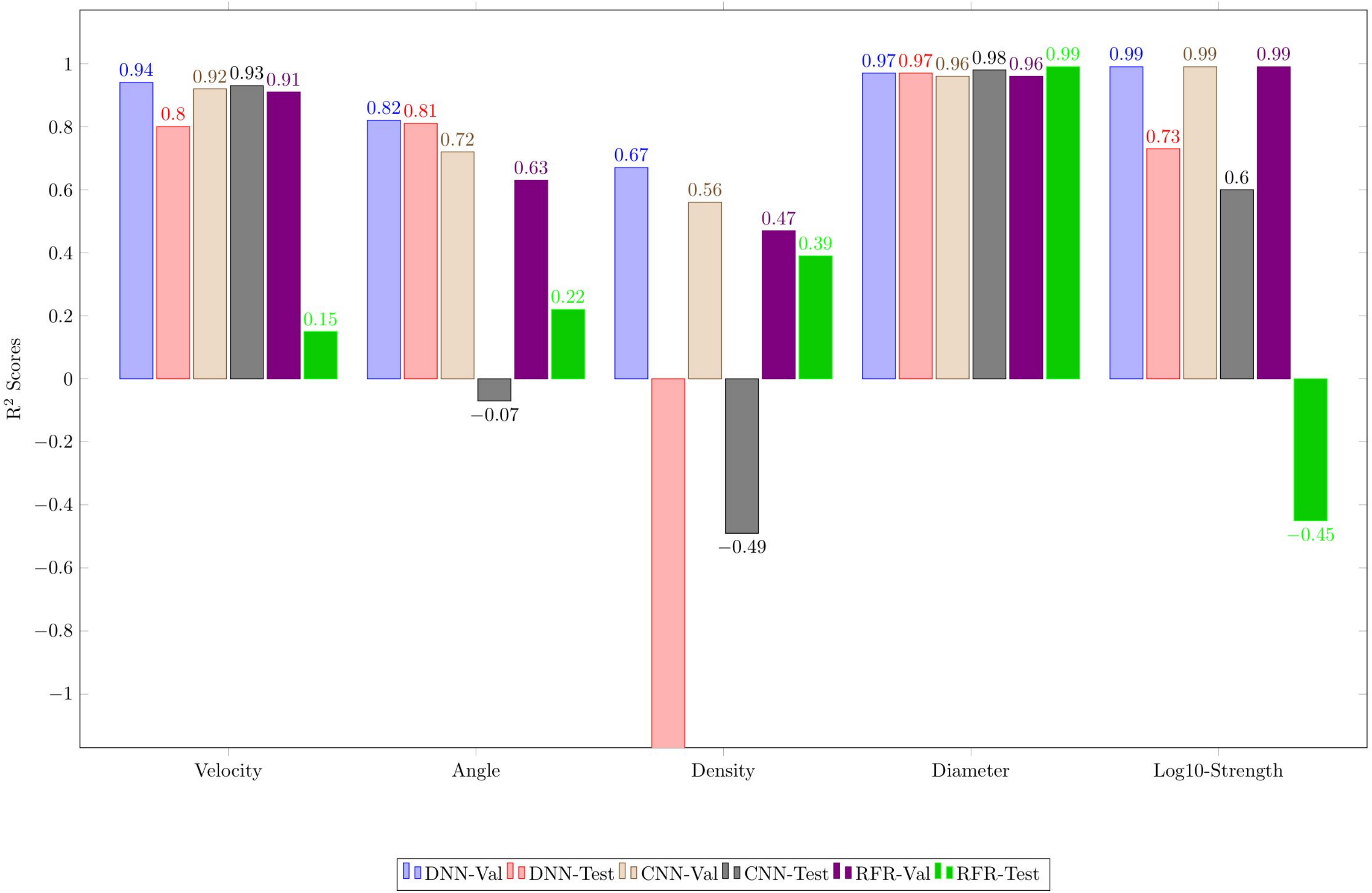


Random Forest Regression



Results

Statistical Comparison of Learning Models Using the Validation and Real Test Data Sets



Results: DNN Best at Angle and Strength

Angle				
Test Case	**Actual** (°)	**Prediction** (°)	**Error** (°)	**Percentage Error** (%)
Deep Neural Network				
Lost City	38.0	50.9	12.9	33.9
Benesov	81.0	77.1	-3.90	-4.81
Tagish Lake	17.8	22.4	4.60	25.8
Kosice	60.0	41.1	-18.9	-31.5
Chelyabinsk	18.3	17.9	-0.40	-2.19
Convolutional Neural Network				
Lost City	38.0	57.0	19.0	50.0
Benesov	81.0	78.3	-2.70	-3.33
Tagish Lake	17.8	68.9	51.1	287
Kosice	60.0	74.5	14.5	24.2
Chelyabinsk	18.3	23.5	5.20	28.4
Random Forest Regression				
Lost City	38.0	53.5	15.5	40.8
Benesov	81.0	51.5	-29.5	-36.4
Tagish Lake	17.8	19.7	1.90	10.7
Kosice	60.0	27.0	-33.0	-55.0
Chelyabinsk	18.3	29.1	10.8	59.0

Aerodynamic Strength				
Test Case	**Actual** (kPa)	**Prediction** (kPa)	**Error** (kPa)	**Percentage Error** (%)
Deep Neural Network				
Lost City	0.25	4.22	3.97	1,590
Benesov	20.0	46.1	26.1	131
Tagish Lake	0.50	1.91	1.41	282
Kosice	2.00	2.97	0.97	48.5
Chelyabinsk	600	1030	430	71.7
Convolutional Neural Network				
Lost City	0.25	6.64	6.39	2,560
Benesov	20.0	32.9	12.9	64.5
Tagish Lake	0.50	4.49	3.99	798
Kosice	2.00	2.61	0.61	30.5
Chelyabinsk	600	679	79.0	13.2
Random Forest Regression				
Lost City	0.25	257	257	103,000
Benesov	20.0	94.4	74.4	372
Tagish Lake	0.50	2.33	1.83	366
Kosice	2.00	19.2	17.2	860
Chelyabinsk	600	634	34.0	5.67

Results: CNN Best at Velocity

Velocity				
Test Case	**Actual** (km/s)	**Prediction** (km/s)	**Error** (km/s)	**Percentage Error** (%)
Deep Neural Network				
Lost City	14.2	13.7	-0.50	-3.52
Benesov	21.5	23.4	1.90	8.84
Tagish Lake	15.8	16.8	1.00	6.33
Kosice	15.0	14.2	-0.80	-5.33
Chelyabinsk	19.2	20.7	1.50	7.81
Convolutional Neural Network				
Lost City	14.2	13.7	-0.50	-3.52
Benesov	21.5	20.0	-1.50	-6.98
Tagish Lake	15.8	16.3	0.50	3.16
Kosice	15.0	15.0	0.00	0.00
Chelyabinsk	19.2	19.2	0.00	0.00
Random Forest Regression				
Lost City	14.2	11.9	-2.30	-16.2
Benesov	21.5	22.4	0.90	4.19
Tagish Lake	15.8	20.7	4.90	31.0
Kosice	15.0	15.1	0.10	0.67
Chelyabinsk	19.2	20.8	1.60	8.33

Results: RF Best at Diameter and Density

Diameter				
Test Case	Actual (m)	Prediction (m)	Error (m)	Percentage Error (%)
Deep Neural Network				
Lost City	0.45	1.47	1.02	227
Benesov	1.35	0.83	-0.52	-38.5
Tagish Lake	4.50	5.56	1.06	23.6
Kosice	1.39	2.44	1.05	75.5
Chelyabinsk	19.8	17.5	-2.30	-11.6
Convolutional Neural Network				
Lost City	0.45	0.65	0.20	44.4
Benesov	1.35	1.89	0.54	21.5
Tagish Lake	4.50	5.01	0.51	11.3
Kosice	1.39	1.81	0.42	30.2
Chelyabinsk	19.8	21.8	2.00	10.1
Random Forest Regression				
Lost City	0.45	0.52	0.07	15.6
Benesov	1.35	1.64	0.29	40.0
Tagish Lake	4.50	3.31	-1.19	-26.4
Kosice	1.39	1.51	0.12	8.63
Chelyabinsk	19.8	19.3	-0.50	-2.53

Density				
Test Case	Actual (g/cm^3)	Prediction (g/cm^3)	Error (g/cm^3)	Percentage Error (%)
Deep Neural Network				
Lost City	3.40	1.54	-1.86	-54.7
Benesov	3.20	3.58	0.38	11.9
Tagish Lake	1.64	1.93	0.29	17.7
Kosice	2.50	1.54	-0.96	-38.4
Chelyabinsk	2.50	3.48	0.98	39.2
Convolutional Neural Network				
Lost City	3.40	2.01	-1.39	-40.9
Benesov	3.20	2.94	-0.26	-8.13
Tagish Lake	1.64	1.48	-0.16	-9.76
Kosice	2.50	3.32	0.82	32.8
Chelyabinsk	2.50	2.07	-0.43	-17.2
Random Forest Regression				
Lost City	3.40	3.42	0.02	0.59
Benesov	3.20	3.47	0.27	8.44
Tagish Lake	1.64	1.76	0.12	7.32
Kosice	2.50	3.53	1.03	41.2
Chelyabinsk	2.50	2.31	-0.19	-7.60

Discussion

- Range of parameters used in training set is limited but representative of asteroid population.
- All the parameters were inferred within modeled/published values.
- R2 scores imply that some parameters are not well constrained.
- Training on synthetic curves leads to good generalizability in the models, if all data points are present.

Diameter	Density	Velocity	Angle wrt Horizontal
2 - 40%	1-41%	0-7%	2-34%
Error from actual model input value	Error from actual model input value	Error from actual model input value	Error from actual model input value

Discussion

- Range of parameters used in training set is limited but representative of asteroid population.
- All the parameters were inferred within modeled/published values.
- R2 scores imply that some parameters are not well constrained.
- Training on synthetic curves leads to good generalizability in the models, if all data points are present.

Diameter	Density	Velocity	Angle wrt Horizontal
18.6%	13.0%	2.73%	19.6%
Mean Absolute Percentage Error	Mean Absolute Percentage Error	Mean Absolute Percentage Error	Mean Absolute Percentage Error

Conclusions & Future Work

Approaches using regression models are much more faster and reliable than GA.
Approaches using regression models are much more usable than GA because it does not require previous knowledge of velocity or entry angle.
Transforming feature space was most important pre-processing approach toward good results.

01 | Apply models to 10 potential other curves.

02 | Use sensor fusion to develop more robust systems.

03 | Develop methodology for incomplete data.

04 | Train models on synthetically generated light curves directly.